

Институт биологии южных морей имени А. О. Ковалевского РАН

PONTUS EUXINUS
ПОНТ ЭВКСИНСКИЙ : XII



ПОНТ ЭВКСИНСКИЙ – 2021

XII Всероссийская научно-практическая конференция молодых учёных с международным участием по проблемам водных экосистем, посвященная 150-летию Севастопольской биологической станции – ФИЦ «Институт биологии южных морей имени А. О. Ковалевского РАН»

Материалы конференции

Севастополь, 20–24 сентября 2021 г.

Севастополь
ФИЦ ИнБЮМ
2021

характерен для пляжей этой части Балтики. Однократных летних съёмок недостаточно для составления полной картины загрязнения пляжей; нужно проводить более частые исследования (раз в сезон, а также до и после штормов и уборок), чтобы иметь возможность более детально отслеживать зависимость распределения микромусора от разных воздействий.

Список литературы

1. Ershova A. A., Eremina T. R., Chubarenko I. P., Esiukova E. E. Marine Litter in the Russian Gulf of Finland and South-East Baltic: Application of Different Methods of Beach Sand Sampling // The Handbook of Environmental Chemistry. Berlin ; Heidelberg : Springer, 2021. P. 1–25. https://doi.org/10.1007/698_2021_746.
2. Haseler M., Schernewski G., Balciunas A., Sabaliauskaite V. Monitoring methods for large micro- and meso-litter and applications at Baltic beaches // Journal of Coastal Conservation. 2018. Vol. 22. P. 27–50. <https://doi.org/10.1007/s11852-017-0497-5>

ОСОБЕННОСТИ СБОРКИ МИТОХОНДРИАЛЬНОГО ГЕНОМА НА ОСНОВЕ КОРОТКИХ ПАРНОКОНЦЕВЫХ ПРОЧТЕНИЙ

Мегер Я. В.¹, Водясова Е. А.^{1,2}, Дмитриева Е. В.², Челебиева Э. С.², Шихат О. В.², Лантушенко А. О.¹

¹ФГАОУ ВО «Севастопольский государственный университет», г. Севастополь

²Институт биологии южных морей имени А. О. Ковалевского РАН, г. Севастополь

Ключевые слова: митохондриальный геном, *Spades*, *Mira*, *MitoZ*, алгоритм *de novo* сборки, парноконцевые прочтения

Вследствие высокой производительности NGS-секвенирования и стоимости получения прочтений, в настоящее время все чаще применяется данный метод секвенирования для получения последовательности полного митохондриального генома. Такой подход заключается в секвенировании со слабым покрытием и за счет перепредставленности последовательностей митохондриального генома происходит его сборка. Однако сам процесс сборки содержит в себе ряд нетривиальных проблем, одна из которых – это разрешение тандемных последовательностей при использовании коротких прочтений.

В данном исследовании были использованы парно-концевые прочтения длиной 100-150 п.н. общим объемом 2.23 Gb. полученных с помощью NGS-секвенирования ДНК *Ligophorus vanbenedenii* на платформе Illumina NextSeq 500. Оценка качества прочтений проводилась с помощью fastqc, которая показала высокое среднее качество прочтений Q30 - 94.07% прочтений. Фильтрация по качеству и удаление адаптеров из массива прочтений производилась с помощью fastp, в результате 95.57% прочтений прошли фильтрацию. Для сборки использовались следующие сборщики: Spades [1] в режиме plasmid, итеративный сборщик MIRA [2], комплексное решение для сборки, аннотации и оценки покрытия MitoZ [3], а также мульти-к-мерный сборщик Norgal. Результаты работы всех сборщиков анализировались на полноту, равномерность покрытия и наличие кольцевой структуры с помощью модуля аннотации MitoZ.

Решение, предложенное в рамках Spades, обладает исключительной простотой в использовании, малым временем выполнения и встроенным механизмом скаффолдинга. Результаты сборки не столь формализованы и требуют дальнейшей

проработки, картирования и аннотации. Однако среди представленных скафолдов есть область с разрешенными тандемными повторами. Это позволяет рекомендовать Spades в качестве черновой сборки для дальнейшей проработки с применением других алгоритмов.

Из-за необходимости последовательности-затравки для работы сборщика MIRA, использование этого алгоритма может быть затруднительно для малоизученных объектов, для которого нет близкородственного референсного митохондриального генома по причине использования последовательности-затравки слабо связанного с исследуемым объектом, что значительно снижает эффективность каждой итерации работы алгоритма. Это привело к чрезмерному времени выполнения сборки. В результате был получен контиг длиной 10834 п.н. с неравномерным покрытием, что является следствием наличия неразрешенного участка тандемных повторов. Алгоритм используемый в этой программе способен разрешить эту проблему лишь при условии наличия хорошей последовательности-затравки или большей длины прочтений. В качестве такой затравки можно использовать результат черновой сборки другого алгоритма, однако такой способ накапливает ошибки и может привести к искаженным результатам.

Мульти-к-мерный сборщик Norgal имеет простой в обращении набор функции и высокое быстродействие, однако, малая длина полученных контигов не позволяет использовать его для получения полной кольцевой сборки митохондриального генома.

Комплексное решение MitoZ предоставляет наиболее наглядное и полное представление о результатах сборки, в автоматическом режиме проводит картирование прочтений и аннотацию. Включает в себя как быстрореализуемый алгоритм черновой сборки, так и более проработанный механизм доработки. Тем не менее конечная сборка имела не кольцевую структуру и содержала область чрезмерного выравнивания – свидетельство разрешения повторов методом кластеризации. С помощью выравнивания полученных скафолдов после Spades эта область была уточнена, что привело к равномерному распределению выравниваний прочтений и достижения кольцевой структуры.

Таким образом для сборки митохондриального генома на основе коротких прочтений целесообразно использовать MitoZ, уточняя области неравномерного покрытия данными других сборщиков или более длинными прочтениями.

Работа выполнена в рамках гранта РФФИ №20-44-920004, государственной бюджетной темы № 121030100028–0 и внутреннего гранта СевГУ 2020 № 33/06-31.

Список литературы

1. Antipov D., Hartwick N., Shen M., Raiko M., Lapidus A., Pevzner P. PlasmidSPAdes: assembling plasmids from whole genome sequencing data // Bioinformatics. 2016. Vol. 32, iss. 22. P. 3380–3387. <https://doi.org/10.1093/bioinformatics/btw493>
2. Chevreux B., Wetter T., Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information // Computer Science and Biology : Proceedings of the German Conference on Bioinformatics. 1999. Vol. 99. P. 45–56.
3. Meng G., Li Y., Yang C., Liu S. Mito Z. A toolkit for mitochondrial genome assembly, annotation and visualization // BioRxiv. 2018. P. 489955. doi: <https://doi.org/10.1101/489955>